DOCUMENT RESUME

ED 262 817 IR 051 328

TITLE IFLA General Conference, 1985. Division on Special Libraries. Section on Administrative Libraries. Papers.

INSTITUTION International Federation of Library Associations, The Hague (Netherlands).

PUB DATE Aug 85

NOTE 83p.; Papers presented at the International Federation of Library Associations (IFLA) General Conference (51st, Chicago, IL, August 18-24, 1985).

PUB TYPE Viewpoints (120) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC04 Plus Postage.

DESCRIPTORS Automatic Indexing; *Databases; Developing Nations; Foreign Countries; Global Approach; Government Libraries; *Government Publications; Information Dissemination; *Information Processing; *Institutional Administration; International Organizations; Library Associations; Library Services; *Special Libraries; Videotex

IDENTIFIERS Access to Information; India; Nigeria; *Universal Availability of Publications; West Germany

ABSTRACT

Papers on administrative libraries which were presented at the 1985 International Federation of Library Associations (IFLA) conference include: (1) "Administrative Libraries and Universal Availability of Information: Comments from a Developing Country" (Diekolola Awosika, Nigeria); (2) "Automatic Indexing and the Dissemination of Information" (Wolfgang von Keitz, West Germany); (3) "Libraries and the Universal Availability of Information: Is Videotex the Missing Link in Information Services? (International Development, Examples and Perspectives)" (Gary L. Harris, West Germany); (4) "Administrative Libraries and Universal Availability of Information with Special Reference to India" (M. K. Jain, India); and (5) "Literature Data Banks for Institutional Libraries. A Critical Comparison between Conventional Information Media and the Possibilities Offered by EDP (Electronic Data Processing)-Supported Legal Information Systems in Europe" (Gerhard Kruse, West Germany). (THC)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

I F L A 51st Council

and General Conference Chicago 1985

Division: Special Libraries

Section: Administrative Libraries

57-ADM-1-E

ED262817

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

# ADMINISTRATIVE LIBRARIES AND UNIVERSAL AVAILABILITY OF INFORMATION: COMMENTS FROM A DEVELOPING COUNTRY

by

Diekolola Awosika
Administrative Staff College of Nigeria
Topo, Badagry.

IR051328

BEST COPY AVAILABLE

*For internal use only:*

To be brought to meeting no:

162 (172 Conf.)

Remarks:

to be presented orally ☒

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A.L. Van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# ADMINISTRATIVE LIBRARIES AND UNIVERSAL AVAILABILITY OF INFORMATION: COMMENTS FROM A DEVELOPING COUNTRY

## Background

Most of the present developing nations were at one time or the other colonies of the developed nations of today. During colonial period the main interest of the colonial masters was in keeping law and order so as to be able to carry out peaceful trade and their missionary work.

Their information requirements related only to orders, rules and regulations as were dictated and sent down by their home governments.

The colonial civil servants were not involved in planning policy and decision making as such, and so their information needs were limited.

However after independence, the situation changed, plans had to be made, decisions had to be taken and policies had to be executed.

The citizens themselves became politically conscious; they need education, they need to know the program of the government, how they are ruled etc. All these lead to the search for information both by the government and the governed. Therefore it became necessary to establish and develop sources and materials from which these informations can be obtained. These requirements gave birth to the development of what we now refer to as libraries in the ministries, departments, research institutes and government parastatals. These collectively are what I will in this paper call the administrative libraries.

Administrative libraries will be those special libraries established to provide management information to decision and policy makers whether in the government or in any organisation for the achievement of its objectives.

The Universal availability of information in relation to administrative libraries is the widest possible scheme for the availability of information in whatever form for administrators, decision makers, opinion moulders, planners and researchers. This group of people need "administrative information" for the effective performance of their job.

Asiedu in 1975 described administrative information as "a meaningful statement about how men, money, materials and machines are managed to achieve the objectives of an organisation"[1] but I shall here put it simply that administrative information is that relevant knowledge required and necessary for the achievement of an organisational objectives be it government or any institution.

Information is an essential element in economic, social, educational, political development of a person and a nation. Therefore the successful working of an information network requires a positive action and the removal of barriers at all levels from the local to international and at all stages from the

publication of new materials to the retention of last copies. Thus it will involve all those connected with the production, storage, dissemination and use of information from the authors, publishers, booksellers, librarians and archivists, documentalists and users to the ultimate beneficiary.

The main sources of administrative information in a country are the government documents which are in form reports, surveys, legislations and Development plans. Therefore the availability of indigenous publications will be a fundamental principle of universal availability of information in any country, that is each country must be able to make its own publications available both to its inhabitants and to other countries. No country can expect to obtain its own publications from other countries. The surest place for obtaining a publication is from the country of publication.

With this fact in mind Nigeria and Africa on the whole has taken some steps to advance the objective of the Universal availability of Publication as promoted by UNESCO, IFLA, British Library Association etc.

Problems facing Universal Availability of Information in developing countries

As much as the developing countries are aware of the need to develop administrative information all efforts have been beset with the following problem.

Publishing

Most information from the government departments are not published and where they are published the number of copies printed is so few that copies are only available within the system.

The publication of information from government departments is handicapped by so many problems. In the first place, publishing is not taken as a serious responsibility as a result most of the useful materials in the government departments are not published. Where the documents are published the numbers produced are so limited and sometimes tagged "Secret" so that they are only available to a few. This situation has deteriorated in the last few years when there had been shortage of paper and other printing materials in most countries. The situation does not seem to be improving either. Part of the problem is that the government departments are not interested in sales promotion so that the marketing and distribution of published documents are so poor that in most cases very few people are aware of what has been published and where they can be obtained. All these are barriers to the availability of information especially when it is realised that most government information are obtained from documents produced in the government departments.

The government of the developing countries should make desperate efforts to publish information emanating from the government departments and arrange for

wider distribution. Furthermore there should be set down criteria for the classification of documents as "confidential" or "secret".

Acquisition

The acquisition of documents especially government publications is the greatest problem of the librarian. When there is no check list it becomes very difficult for any librarian to do this selection yet most governments in the developing countries do fail to publish a list of their publications and when the lists are available they are usually not current as a result the librarian does not know what to acquire. The government documents are not obtained in the open market, the marketing and distribution is therefore closed to the general books stores and booksellers that could give easy access. The government bookshops will only sell on prepayment; a practice which most libraries do not favour especially since there is the risk of loosing the documents in the post or of their not been sent at all when the documents are already out of stock. Most government documents are soon out of stock, almost immediately they are published because limited copies are produced and it is not usual for government to have reprints.

In most cases, publishers only pay lip service to the legal deposit laws which are rarely enforced, in fact some legal deposit libraries have had to police some publishing houses in order to collect copies of latest publications. They usually detail a librarian to go round government publishing houses to collect documents.

If the acquisition of government documents must be eased for the administrative libraries, the government should where this has not been done give legal backing to the acquisition of the materials. Gifts and exchanges should be encouraged and organised among the libraries so that each of the libraries may have a comprehensive stock.

Cooperative Scheme

Documents are very rare materials because of the difficulties encountered in their acquisition therefore most libraries are reluctant to give them out on Interlibrary loan. This creates a problem in cooperative scheme and effects the Universal Availability of Information.

In most cases, cooperative schemes are not well planned and programmed before they take off hence many libraries enter into it half-heartedly and only lend materials on informal basis. Finance is another problem that causes a set back in a cooperative scheme. Where a library is willing to send a photocopy the question arises as to who should be responsible for the cost of the photocopy. Is it the lending library or the borrowing library? This cost could be exhorbitant when there is a heavy demand on a particular library. Thus the library will need extra

funding for photocopying machines their maintenance and operating materials. The cost of maintaining photocopying machines in some developing countries like Nigeria is very colossal. The frequent breakdown of the machines is caused by electricity failures and lack of spare parts.

Organisation and storage of the materials

The contents of the documents may not be specific on any particular subject, then the problem of location in the general classification schemes become difficult. In Nigeria, the National Library of Nigeria developed a special classification scheme for government documents. Another problem is in the form in which these documents are produced; usually they are flimsy and in paper backs which makes storage, and handling difficult. They are unable to stand several uses before they fall apart. For storage, they cannot stand on the shelves and have to be kept in most cases in pamphlet boxes. The best way to keep documents for prosperity is either to have them bound in hard cover and subsequently in micro form.

Despite the above problems Nigeria and other African countries have made the following efforts for the availability of information.

Establishment of Administrative Libraries

In most developing countries, most government departments and ministries have departmental libraries which serve as information centre for the public servants. These libraries play very prominent and important role in the flow and dissemination of administrative information within the rank and file of the Public Service. They acquire and store published, unpublished and often restricted materials ranging from annual reports, technical reports, statistics to administrative and policy statement within their respective departments.

Ghana Situation

It is in appreciation of the value of administrative information in the effective performance of government functionaries that in 1974, Asiedu[2] proposed the establishment of a Central Management Library and Documentation Services Department in Ghana. The CMLDA was to be a "special department within the public service system forming an organic part of the policy design and formulation process of governmental operation, contributing towards quality, efficiency and economy".

The main function of the CMLDS was to collect, store and disseminate administrative materials in all forms to meet the information and research needs of the entire governmental administrative machinery: and to support the policy of the Central Government as follows:

(i) coordinate all the activities of the departmental libraries of the government;

(ii) provide reference and documentation services;

(iii) provide current awareness services for the top management personnel of the government;

(iv) exploit the documents thus collected by disseminating their contents to the practising administrator;

(v) provide indexing, abstracting and translation services on the various subjects of interest to the administration.

This is a laudable scheme which if implemented will serve as a nucleus of a network for the dissemination of administrative information in that country. Thus laying a foundation for the universal availability of information.

The Ghana Institute of Management and Public Administration (GIMPA) has a well established library housing the National Documentation Centre for Management and maintains close links with administrative staff colleges and Institute of Public Administration in many countries. The library is also the National Focal point for the African Network in Administrative Information which will be described later in this paper.

Nigeria Situation

Like Ghana, Nigeria has made some progress in her attempt to further the objectives of the UAP. The Federal integrated library service system is a collection of the government libraries in the ministries and department. Each library in a department is responsible for the acquisition, storage and making available all materials published in their respective ministries, and other materials relevant to the functions of the department.

In recent times the government of Nigeria set up many training institutes for the administrators and policy makers in the Public Service, and the armed forces. To mention a few, these are the Nigerian Institute of Policy and Strategic Studies, Nigerian Institute of International Affairs, The Centre for Management Development and the Administrative Staff College of Nigeria. The common function of all these institutions is in the establishment of a library. This provision is specifically stated in the statutes establishing each of the institutes.

As a further step to making maximum use of available information, libraries in Nigeria have organised a cooperative acquisition and lending scheme under the aegis of the National Library of Nigeria. The scheme is to maximise financial resources of all libraries in the country and to provide as comprehensive a collection as possible nationally for the benefit of all. The country is divided into zones with each zone to acquire the subjects allocated to them. Activities of zones are to be coordinated at the zonal headquarters under the zonal coordinator. Acquisition lists of all participating libraries are to be sent to zonal headquarters where they will be sent to the National Secretariat for final compilation of the whole scheme and then distribute. The important point here is that all the administrative libraries participate in this cooperative scheme,

thus making it possible to spread out the information in each library.

In 1981, a circular was sent round to all Federal and State Ministries and Parastatals stating the need to keep in a central place, copies of all government documents pertaining to past and present administrations with particular reference to reports of different commissions, government white papers, salary reviews etc. The Administrative Staff College of Nigeria was considered suited for the storage and keeping of such documents. Laudable as this proposal is, the response to this suggestion has been very poor.

The Nigerian Institute of International Affairs is now developing an oral documentation centre where recorded interviews with all past and present public figures, retired civil servants, executives of corporations and firms and politicians are stored. The topical issues raised in the interviews covered social, economic, political and development plans and policies in the country.

In the continent, within the last decade efforts have been made to launch cooperative administrative information network schemes. Thus in 1975 the African Training and Research Centre in Administration for Development (CAFRAD) convened a meeting attended by representative of seventeen African countries. The object of the meeting which was held in Accra, Ghana was to consider the lack of adequate and reliable administrative information for making policy and administrative decisions in the continent. The result of this meeting was the establishment of ANAI - African Network in administrative Information.

ANAI (African Network in Administrative Information)

The African Training Research Centre in Administration for Development (CAFRAD) recognised the lack of adequate and reliable administrative information for making policy and administrative decisions in African countries. It has as its objectives to promote the development of its library and documentation centre as an African administrative clearing-house and using a network approach, the effective sharing, transfer organization and dissemination of administrative information and training material throughout the continent. This clearing house function is to accumulate at CAFRAD all information on administrative and managerial development in Africa.[3] Therefore a seminar for African Librarians and Documentalists funded mainly by Commonwealth Fund for Technical Cooperation was held in Accra, Ghana in 1975. At the seminar which was attended by representatives mainly from Anglophone African countries it was realised that the problem of developing countries was not due to lack of material resources but lack of efforts to make the available information reach the users and for the users themselves to appreciate the existence and usefulness of these materials. Hence a resolution was adopted for the setting up of the African Network in Administrative Information (ANAI) whose main objective was to make available to decision-makers and

opinion moulders, administrators, planners, researchers etc. up-to-date relevant and accurate information in the areas of development planning process.[4] A similar seminar was organised for the Francophone countries at Tangier in 1978.

In November 1983 another meeting of representatives from participating countries was held in Dakar and the objectives of the meeting among others were:

(a) to formalise the institution-building aspects of ANAI;

(b) to inaugurate the Network Inter-Library cooperation Scheme;

(c) to consider the definitions of the substantive discipline of development administration.

Again at the seminar it was established that the problem of African countries in Information for National Development is not so much lack of information but

(a) in the utilization of the available information resources;

(b) the organisation and coordination of the resources for effective utilization;

(c) the orientation of the users for an appreciation of its usefulness.[5]

PADIS Project (Pan African Documentation and Information System for Social and Economic Development)

Recognising the need for increasing self-reliance among African States in order to accelerate the wide-spread development of capabilities in the various fields of technical, social and economic development, the United Nations Economic Commission for Africa (UNECA) in 1980 set up the PADIS Project. The main operational objectives are "to provide ready access to information for policy makers, technicians, planners, financiers and others engaged in the economic and social development of African States".[6, 10]

Among the efforts which the ECA hope to make to fullfil the objectives were the setting up of a Central Coordinating body and the creation of a data base at the Headquarters.

The ECA also hopes to assist member states in strengthening their national infrastructures in order to enable them to participate in PADIS. It also plans to establish linkage with international information networks. So far there is no information on the extent to which PADIS has achieved its objectives.

The setting up of the projects mentioned above augurs well for the Universal Availability of Information but the problem is in their prosecution to achieve the desired goals. The problems hindering the proper execution of the projects are important issues to which attention should be paid.

An important issue is that meanwhile very few national governments show any commitment to these projects. As a result the national focal points or centres where they do exist at all are given little or no recognition. The non-recognition

of the National Centres results in non-performance by these centres which have no fund to execute the projects. Hence the necessary materials and infrastructure are not available.

There is dearth of information and publicity on the projects so that even the participating centres have little or no knowledge of what the projects are about and what they are supposed to do.

Where seminars are organised the participants are given very short notices and are not properly briefed before the meetings. Thus deliberations at the meeting showed that most delegates have no mandate to speak for their governments and after the seminar do not know the right authority to whom they should report.

There is no continuity in that the staff in the ministries are changed too often. Once a staff on the project is transferred his successor may not be aware of the existence of such a project or he may not be as interested thus the programme dies a natural death.

It is therefore desirable that the government or each member country gets involved and committed to the projects. A way of achieving this may be to mount seminars on each of the projects at the national levels where ministers, government officials, policy-decision makers and librarians and others responsible for the projects will meet. National Centres for each project should be identified and each of the National Centres should be given extra budget for the execution of the projects.

As stated before a large number of sources of Information in the Administrative Libraries are from Government Publications. These will be in the form of reports, statistic surveys etc. issued by the ministries and government agencies. To be able to acquire these materials, it is necessary for the National centres to be statutorily designated as legal deposit centres for all government publications. Despite these shortcomings "CAFRAD's Documentation Centre is said to have emerged as one of the strongest, best endowed specialised documentation centres in Africa. It is favoutably equipped for a lead role among the network of African documentation centre".[7] So far it has produced:

(1) a draft Thesaurus of Administrative Information descriptors;

(2) a Directory of Administrative Information sources to be updated periodically and

(3) an annotated bibliography of Public Administration in English West Africa.

Conclusion

A statement which has been credited to Bowden is that "the librarianship of developing countries is scattered with tombstones as memorials to the cooperative projects that died or of the corpses that barely cling to life"[8] should this continue to be the case for all projects initiated in the developing countries

in Africa? To avoid the continuation of and debunk this belief it is necessary for the developing countries to review the organisational framework in which they operate. The universal availability of information should start at the grass root level. All administrative libraries within a country should first form a cooperative scheme of acquisition lending, exchanges and documentation so that collectively they can form a nucleus for the successful execution of the ANAI project.

The pilot phase of the project has already taken ten years, yet not concrete effort has been reported on any of the participating countries. The latest report is that "CAFRAD's Library can boast of considerable collection of rare publications on African Administrative structures, systems, practices, reforms etc.".[9] The question is how soon do we set up the machinery to make the information available to all participating countries?

1. Asiedu, E. S. - Organising the flow of Administrative Information in a developing setting: a proposal for the Ghana Public Services. Greenhill Journal of Administration Vol. I No. 3 Oct. - Dec. 1974 pages 39 - 51.

2. Asiedu, Op cit.

3. Boadi, Berizies Y. Regional Library Cooperation for development in Africa, IFLA Journal 10 No. 4 1984 pages 370 - 376.

4. CAFRAD - African Network of Administrative Information: ANAI Experimental (A Three Year Pilot Project) 1985/86 - 1987/88 (unpublished).

5. CAFRAD AIDE Memoir - African Network of Administrative Information (ANAI) unpublished.

6. CAFRAD - Report of the ANAI meeting on African Administrative Information Services on Cooperation (West Africa sub-Region) Nov. 14 - 19 1983 (unpublished).

7. Hadi, Mohammed, M. El
The African Administrative Information and documentation activities of CAFRAD. Unesco Bulletin for Libraries. Vol. 31 No 3 May - June 1977. Item 167.

8. Hadi, Op cit.

9. United Nations, Economic Commission for Africa PADIS Padis/80/6/Rev./

10. see also:
Gehrke, Ulrich Information for Development:

Some Problems of National Co-ordination, Regional Co-operation and International Assistance. Paper given at the 50th IFLA-Council and General Conference Nairobi 1984 (to be published in Inspel 1985).

IFLA 51st Council
and General Conference, Chicago 1985
Division: Special Libraries
Section: Administrative Libraries

58-ADM-2-E

# Automatic Indexing and the Dissemination of Information

by

Wolfgang von Keitz
Universität des Saarlandes

Saarbrücken
FR of Germany

*For internal use only:*
To be brought to meeting no:
162

Remarks:
to be presented orally ☒
for distribution only ☐

## 0. INTRODUCTION

In 1982 about 100.000 periodicals are published all over the world. And serious estimations talk about more than 5 million specialized publications per year (essays, books). Our libraries have the important task to store this universe of information and to make it available to a growing demand. This steadily growth of published information and the continous increase of information demand calls for new and automatic library systems.

Traditionally, access to information, stored in a library, is made by a collection of file-cards. These cards are alphabetically arranged. The user has to know an author or a controlled subject to start a manual information search. In some well-established libraries people can use an information retrieval system. And it is obvious that the convenience and usefulness of these systems make the traditional way of searching information in a library obsolete.

People realize that electronic search tools are userfriendly because the search can be done very quickly. A most useful improvement allowed by only electronic retrieval systems could be obtained by a sophisticated subject search. The user must only know the subject under which he is searching information and the system will show him all the documents of a specific subject.

It is evident that the quality of the search depends on the quality of the index of the retrieval system. The task of an index is to identify all the relevant documents to a query. The indexer's or librarian's task then is to design indexes fulfilling this task.

With regard to the significance of the index what the quality of an information retrieval system is concerned, it is not surprising that the methods in designing a good index have been in

controversial discussion. The importance of the indexes seems quite clear. Nevertheless, anybody working in this field knows the problems involved in the determination of indexes, i.e. "that the selection of information pertinent to a particular problem or interest, from all the previously recorded information available, involves problems of decision making base on less than the full content or text of each records being searched " (STEVENS; 1965, S.2).

The task known as "indexing" involves

1. choosing terms or concepts that will serve to identify a document for later retrieval

   and

2. assigning to each term a weight, or value reflecting its presumed importance for purposes of content identification (SALTON, 1983, S.59).

Term weighting procedures are exhaustively discussed by SALTON in his "Introduction to modern Information Retrieval" (SALTON, 1983). Therefore we can concentrate on the first part of the task.

Choosing terms by computerized means is called "automatic indexing". The growing use of computers in compiling indexes must be seen on the basis of the still increasing rate of production of scientific literature and reports and the restricted manpower in libraries. But in the mainstream the computer is still "ante portas" of the libraries.

At the present time two principal indexing strategies are used (SALTON, 1982, S.66):

First and in the most situations, the indexing is performed manually by trained indexers, who assign to each document terms that may by taken either from a freely chosen or a controlled vocabulary. Typically the content of a document is represented by 10 terms.

Second and often used in combination with the first strategy, indexing is performed by a computer, assigning all the words of the document except for a few function words. In this case the content of the primary document is represented by a collection of terms extracted from the title and the abstract of the document.

The quality of the manual indexing is dependant on the experience and the specific scientific background of the indexer. On the opposite the qualitiy of the automatic indexing system is dependant on the linguistic ability of processing text words and the statistical ability to evaluate these words with respect to their document identification quality.

On this background the well-known controversy "indexing by machine" versus "indexing by man" is rather academic. If the automatic indexing systems performed the task of choosing identifiers to a document as well as an indexer does and if the system was faster and cheaper than the indexer anybody would buy an automatic indexing system. For obvious economic reasons we need automatic indexing systems to solve the problems of sheer amount in terms of manpower requirements and time necessary to produce indexes.

Unfortunately there is no computershop where librarians can have a look at the various types of automatic indexing systems. To choose a system, which fits one's need best is really a problem.

It is true, that at the moment a system of an utmost quality does not exist. Therefore, in view of the further necessity of

automatic indexing systems a summary report at this time should answer two questions, to give a decision support for the librarian.

The first question is: What can by properly done by machine? And second: What parts of the indexing task remain to be done by man if the indexing task is actually performed by machine?

In the following, four automatic indexing systems will be described, which are in practical use as pilot systems or as commercial systems in West-Germany. We will not take into account experimental systems, like the very famous SMART-System developped by Gerald SALTON, nor semi-automatic systems like 'PRECIS' designed by the BRITISH LIBRARY.

We will introduce the system AIR, that simulates manual indexing by machine, as well as a full text indexing system based on the extraction of text words, the dictionary-based indexing system PASSAT and the dictionary and rule based indexing system CTX. We will provide an example for each system to illustrate the indexing results of each system. Finally relations are drawn to automatic translation.

## I. AIR - SIMULATION OF MANUAL INDEXING

Since 1978 an automatic indexing system has been developed at the Technische Hochschule Darmstadt under the guidance of G. LUSTIG. This system shows a completely different strategy than the other systems to be introduced in the following (LUSTIG, 1984).

Documents newly added to a retrieval system are indexed by AIR (Automatic Indexing and Retrieval System) with the help of stochastic functions. These functions are computed as a

relationship between text words of a document and the previously manually assigned identifiers. Each document is represented by an identifier that originates from a given indexing vocabulary. Such, a presupposition of the application of AIR is the developmentof a suitable indexing vocabulary (i.e. thesaurus). The single terms of the indexing vocabulary are then related by means of a descripton vector with the text words of already manually indexed documents. The description vector contains detailed information about occurence, relevance, distribution frequency and document type. The result is a stochastic descriptor-document relation which builds the starting point of automatic indexing.

The structure and the way of functioning of this approach are described in a more detailed way by KNORZ (KNORZ, 1983). However, it is sufficent for us to know that this method assigns to a document approximately the same identifiers as an indexer does. The application of this method is to be recommended wherever already satisfying manual indexing systems have been applied but should be now substituted by an automatic indexing system because of an increasing amount of incoming documents.

On the other hand, if we have to cope with an indexing environment in which not yet controlled term lists or thesauri are available methods are preferred which use the actual text words of a document for the purpose of automatic indexing.

## II. FULL TEXT INDEXING

During the last twenty years refined automatic indexing techniques have been developed, in order to assign identifiers directly out of a document text. These techniques use sophisticated text analysis and identifier weighting methods. But nevertheless one of the mostly used methods is full text indexing.

First of all the process of indexing must start with the identification of all individual words that constitute the document. In the first indexing step the high-frequency function words are to be eliminated. This procedure condenses the text to 30 or 40 percent of text words and can be done by a simple dictionary checking. The dictionary or stop word list may include 10 up to 200 words. In the next step the remaining text words are alphabetically arranged in an index file.

Table 1: Document, Database INFODATA, Gesellschaft für Information und Dokumentation, Frankfurt, W-Germany

| | |
|---|---|
| INFO84001521 Document | |
| COPYRIGHT | Copyright: INFODATA. GID-Informationszentrum. Frankfurt/M |
| STANDORT | In: ZH1 83-0492 |
| AUTOR | Heilmann W. (Integrata, Tübingen, DE) |
| TITEL | Organisatorische Gestaltung von informationstechnisch gestützten Heimarbeitsplätzen für Programmierer (Teleprogrammierung) |
| KONF | Gesellschaft für Informatik (GI), Bonn, DE; Arbeitsgemeinschaft Datenverarbeitung (ADV), Wien, AT: Büroinformation und -kommunikationssysteme, Wien, AT: 1982.09.30-1982.10. |
| QUELLE | In: Hansen, H.R. (Hrsg.): Gesellschaft für Informatik (GI Bonn,DE; Arbeitsgemeinschaft für Datenverarbeitung (ADV), Wien, AT: Büroinformations- und -kommunikationssysteme. Anwendergespräch, Wirtschaftsuniversität Wien, 30.09. bis 01.10.1982. Berlin, DE: Springer (1982) S. 243-263, 4 Abb 31 Lit., de- Betriebs- und Wirtschaftsinformatik 2, ISBN 3-540-11599-4 |
| TEXT | Befasst sich mit den Ergebnissen einer "Studie über Auswahl, Eignung und Auswirkungen von informationstechnisch ausgestalteten Heimarbeitsplätzen". Beschreibt ein im Zusammenhang mit dieser Studie durchgeführtes Pilotprojekt, das die Errichtung von Telearbeitsplätzen für Programmier und die qualitative und quantitative Analyse der technischen und organisatorischen Bedingungen und der wirtschaftlichen und sozialen Auswirkungen der "Teleprogrammierung" zur Aufgabe hatte. |
| DESK | Tischcomputer; Telekommunikation; Arbeitsplatzgestaltung; Wirkungsforschung; Informationstechnologie; Programmierun Arbeitszeit; Arbeitsablauf; Gesellschaft; Pilotanwendung |
| REF-NR | Z 8408 1846 |

BEST COPY AVAILABLE

This index file is the index part of the so called inverted file system. That is, an inverted file system consists of a document file and one or more directories known as the "inverted index". The inverted index contains the text words, which are determined by full text indexing. For each term an associated list of documents reference numbers is included in the index.

Table 1 shows a document stored in the database INFODATA which is produced by the Deutsche Gesellschaft für Information und Documentation (GID). To retrieve this document the searcher may use the controlled vocabulary index (CVF) assigned by an indexer and containing identifiers like TISCHCOMPUTER or TELEKOMMUNIKATION (Table 2).

Table 2: Identifiers, Controlled vocabulary by an Indexer

Arbeitsablauf
Arbeitsplatzgestaltung
Arbeitszeit
Gesellschaft
Informationstechnologie
Pilotanwendung
Programmierung
Telekommunikation
Tischcomputer
Wirkungsforschung



BEST COPY AVAILABLE

Or he may use the full text index (FTI) (Table 3); or he may combine CVI and FTI.

Table 3: Inverted File Full Text Index

Analyse
Aufgabe
Ausgestaltetan
Auswahl
Auswirkungen
Auswirkungen
Bedingungen
Befasst
Beschreibt
Durchgefuehrtes
Eignung
Ergebnissen
Errichtung
Gestaltung
Gestuetzten
Hatte
Heimarbeitsplaetzen
Heimarbeitsplaetzen
Informationstechnisch
Informationstechnisch
Organisatorische
Organisatorischen
Pilotprojekt
Programmierer
Programmierer
Qualitative
Quantitative
Sozialen
Studie
Studie
Technischen
Telearbeitsplaetzen
Teleprogrammierung
Teleprogrammierung
Ueber
Wirtschaftlichen
Zusammenhang

BEST COPY AVAILABLE

A short comparison between hand made controlled vocabulary index and the automatic made full text index shows that the correspondence between the two indexes is very low (Table 4).

Table 4: Controlled Vocabulary Index VS. Full Text Index

| Controlled Vocabulary Index | Full text Index |
|---|---|
| Arbeitsablauf | - |
| Arbeitsplatzgestaltung | Gestaltung |
| - | Heimarbeitsplaetzen |
| Arbeitszeit | - |
| Gesellschaft | sozialen |
| Informationstechnologie | informationstechnisch |
| Pilotanwendung | Pilotprojekt |
| Programmierung | Programmierer |
| Telekommunikation | Teleprogrammierung |
| Tischcomputer | - |
| Wirkungsforschung | - |
| | Analyse |
| | Aufgabe |
| | Ausgestalten |
| | Auswahl |
| | Auswirkungen |
| | . |
| | . |
| | . |

BEST COPY AVAILABLE

First of all, the full text index contains more identifiers than the controlled vocabulary index. Second FTI contains all words of the text in its natural occurence; CVI contains only noun words. Third, CVI-Identifiers represent sometimes two or more FTI-Identifiers (CVI: Arbeitsplatzgestaltung, FTI: Gestaltung, Heimarbeitsplätzen). And last, CVI represents the concept of the document and FTI represents the text of the document. As a summary you may say CVI is rather accurat in describing a document, but FTI is more exhaustive.

As you may remember, our first task in automatic indexing has been "to choose terms or concepts that will serve to identify a document for later retrieval". The full text indexing method fulfills it's task. High frequency function words are eliminated and text words are chosen as identifiers. Full text indexing has been often criticized. But in spite of that criticism it works.

If you are a critic of full text indexing and if you want an improved automatic indexing system, you need a system which removes word suffixes and reduces the original text words to word stems.

## III. PASSAT - A dictionary based automatic indexing system

PASSAT is a system developed and marketed by SIEMENS AG, München for editing documents and analyzing the contents of texts. The characteristic keywords of the text are selected with the aid of a dictionary, the so called "comparative word list" (CWL). With the help of these "comparative word list" entries the text words are reduced to word stems, compounds are divided up into their components, nouns in general are supplemeted by their corresponding verbs and vice versa (derivations), or they are

replaced by synonyms or supernyms. PASSAT also offers foreign-language identifiers. The process of applying text word substitutions provides the opportunity of using a controlled vocabulary (thesaurus).

The actual text analysis and description of the documents requires four steps:

1. alphabetic sorting of text words
2. text word reduction to word stems
3. elimination of stop-words
4. supplements contained in tne "comparative word list" are assigned to the list of identifiers.

The heart of PASSAT is it's dictionary (Table 5).

Table 5: PASSAT dictionary, excerption

```
AA: DV02FFFF:AUSWAERTIGES AMT
AACHEN: DE42F110
AAL: DE07CA10
AARGAU: DE42F110
AARHUS: DE424010
AAS: DE09D610
AB: WK44FFFF
ABAENDER: DV06A610: ABAENDERUNG
ABAENDERBAR: DE41F7FF
ABAENDERUNG: DV08CF17: ABAENDERN
ABBAU: DE076210
ABBAU: DE46A4FF
ABBAUBAR:DE01F7FF:ABBAU
```

An entry of the dictionary is constructed in the following manner:

ABAENDER;DV06A610;ABAENDERUNG

substitute
Bindungsnumber
Suffix number
part of speech
substitution type
identifier marker
word-stem

Table 6 shows a dictionary based indexing by PASSAT.

Table 6: Dictionary Based Index by PASSAT

| | |
|---|---|
| ANALYSE | HEIM |
| AUFGABE | INFORMATION |
| AUSGESTALTEN | INFORMATISIEREN |
| AUSGESTALTUNG | INFORMIEREN |
| AUSWAHL | INFORMIERUNG |
| AUSWIRKUNG | PILOT |
| AUSWIRKEN | |
| BEDINGUNG | ARBEIT |
| BEDINGEN | ARBEITSPLATZ |
| BESCHREIBEN | PROJEKTE |
| BESCHREIBUNG | PROGRAMMIERUNG |
| DURCHFUEHREN | PROGRAMMIEREN |
| DURCHFUEHRUNG | PLATZ |
| EIGNUNG | TELEARBEITSPLATZ |
| EIGNER | HEIMARBEITSPLATZ |
| ERGEBNIS | TELEPROGRAMMIERUNG |
| ERRICHTUNG | PILOTPROJEKT |
| ERRICHTEN | INFORMATIONSTECHNISCH |
| GESTALTUNG | |
| GESTALTEN | |
| STUETZEN | |
| STUETZUNG | |
| RABEN | |
| ORGANISATORISCH | |
| PROGRAMMIEREN | |
| QUALITATIV | |
| QUANTITATIV | |
| SOZIAL | |
| STUDIE | |
| TECHNISCH | |
| WIRTSCHAFTLICH | |
| ZUSAMMENHANG | |

How a comparision shows between full text indexing and dictionary based indexing (Table 7), apart from the reduction of text words to word stems an expansion of the index is achieved (i.e. more identifiers are assigned) with the help of the information contained in the dictionary. Such, AUSGESTALTETEN is reduced to AUSGESTALTEN and supplemented by AUSGESTALTUNG. Hence, HEIMARBEITSPLATZEN is not only generated into HEIMARBEITSPLATZ but also all the meaningful parts of the compound like HEIM, ARBEIT, PLATZ, ARBEITSPLATZ are put into the index.

Table 7: Full Text Index VS. Dictionary Based Index

| Full text indexing | Dictionary based indexing |
|---|---|
| ANALYSE | ANALYSE |
| AUFGABE | AUFGABE |
| AUSGESTALTEN | AUSGESTALTEN |
| - | AUSGESTALTUNG |
| AUSWAHL | AUSWAHL |
| AUSWIRKUNGEN | AUSWIRKUNG |
| . | AUSWIRKEN |
| . | . |
| . | . |
| . | . |
| HEIMARBEITSPLATZEN | HEIM |
| . | ARBEIT |
| . | PLATZ |
| . | ARBEITSPLATZ |
| . | HEIMARBEITSPLATZ |



BEST COPY AVAILABLE

This advantage has it's price, of course. The full text indexing is performed automatically, whereas the dictionary based indexing system has to be looked after by an indexer.

Text words, not included in the dictionary, are listed in a special file. The indexer has to control that there aren't typing mistakes in the document text or that the item is actually not contained in the dictionary. In this case the dictionary is to be complemented.

By means of PASSAT several hundreds of documents per day can be processed. The indexer has to face a slightly different work situation. Instead of assigning identifiers he has to control document texts and to complement dictionaries.

From our short description of PASSAT it becomes evident that automatic indexing systems require continuous work on the dictionary. This work will be more and more an effort the more effective the indexing system is supposed to be.

## III CTX - A dictionary and rule based indexing system

A further improvement in automatic indexing is the CTX-system, developed at the Univesity of Saarland (ZIMMERMNANN, 1982, 1983). CTX additionally to single terms generates complex terms which has the effect of being precisely the inverse of word derivation or word supplements.

Before introducing the method in detail some remarks on the function of CTX and it's history. CTX represents the result of twenty year work at the University of Saarland under guidance of Harald ZIMMERMANN.

The practical application of CTX which is presented here, is based on a development which has it's beginning in 1962. At that time the goal of H. EGGERS, a german pioneer in automatic text processing, was to develop a grammar of German with the help of a computer. In the beginning of the seventies a project was launched, finnacially supported by the Deutsche Forschungsgemeinschaft to develop a text analysis- and translation system. The theoretical framework of this early work has been based on findings of the transformational grammar and parts of value theorie. On the other hand substantial emphasis was on a systematic lexical approach in coding a morpho-syntactical dictionary containing more than 80.000 entries. This text analysis system is one of the main modules of CTX. Leaving out of consideration the round-about-ways and the dead ends following during the basic research works in this area the expenditure of work for the development of the whole system is about 120 - 150 men years.

The realization of an efficient dictionary component and the continuous improvement of thedictionary is one of the most important tasks. Colloquial words are registered in a morpho-syntactical and in a semantic dictionary. Subject-specific words are registered in a special dictionary with semantic relations (thesaurus). Rather than giving a detailed description of the structure of the dictionary we turn to introducing the steps a text takes to indexing with CTX.

CTX is a dictionary and rule based indexing system extracting document identifiers, taking into account syntactic and semantic problems in automatic indexing. CTX has a modular structure. It's application is limited to written German.

The term extracting process consits of six steps:

Step 1: Reduction text word to word stem

| text word | basic form |
|---|---|
| Vorzüge | VORZUG |
| trab | TRETEN |
| trifft ... zu | ZUTREFFEN |

Step 2: Decomposition of composite words

| text word | decomposition in basic form |
|---|---|
| Persönlichkeitsstärke | PERSÖNLICHKEIT |
| | STÄRKE |
| Generalstabschef | GENERALSTAB |
| | GENERAL |
| | STAB |
| | CHEF |

Step 3: Disambiguation of ambigue words

| text words | disambiguation in basic form |
|---|---|
| in der Praxis der | PRAXIS = praktisches Vorgehen ( in contrary "Arztpraxis") |

Step 4: Construction of term phrases

| text words | term phrases in basic form |
|---|---|
| tritt ... in Kraft | IN KRAFT TRETEN |
| personenbezogene, durch das Gesetz geschütze Daten | PERSONENBEZOGENE DATEN DURCH DAS GESETZ GESCHÖTZTE DATEN. |

Step 5: Construction of complex identifiers

| text words | complex identifier in basic form |
|---|---|
| moderne Industriestaaten | INDUSTRIESTAAT<br>MODERNER INDUSTRIESTAAT |

Step 6: Elimination stop-words

A part of the index produced by CTX is shown in table 7. You have to notice that apart from the normalized text words adjektiv-noun relations (ORGANISATORISCHE GESTALTUNG, AUSGESTALTETER HEIMARBEITSPLATZ) are assigned as well as shortened phrases (GESTALTUNG HEIMARBEITSPLATZ G, AUSWAHL EIGNUNG K).

Let us consider the precoordinated index of CTX on a closer perusal. Just like in the examination of the struture of items with several words in the manual precoordination three types of identifiers are brought in use (ZIMMERMANN, 1983, 52ff.):

1. adjective noun relation
   ORGANISATORISCHE GESTALTUNG

2. noun noun relation
   EIGUNUNG DES HEIMARBEITSPLATZES

3. noun preposition attribute relation
   STUDIE ÜBER DIE AUSWIRKUNGEN

These are the presupositions for CTX for generating up to four different noun combinations (complex identifiers). A complex identifier consists of a couple of single indentifiers and special characters.

To list the possible combinations:

1. adjectiv noun relation
   ORGANISATORISCHE GESTALTUNG

2. noun noun genitiv relation
   GESTALTUNG HEIMARBEITSPLATZ G

3. noun noun conjuntion relation
   AUSWAHL EIGNUNG K

4. noun preposition attribute relation
   STUDIE AUSWAHL P

An example of a CTX-Index is schown in Table 8.

Table 8: Single Term and Phrase Index by CTX

```
ORGANISATORISCH
GESTALTUNG
INFORMATIONSTECHNISCH
GESTUETZT
HEIMARBEITSPLATZ
PROGRAMMIERER
ORGANISATORISCHE GESTALTUNG
GESTUETZTER HEIMARBEITSPLATZ
GESTALTUNG HEIMARBEITSPLATZ G
GESTALTUNG PROGRAMMIERER P
TELEPROGRAMMIERUNG
BEFASSEN
ERGEBNIS
STUDIE
AUSWAHL
EIGNUNG
AUSWIRKUNG
INFORMATIONSTECHNISCH
AUSGESTALTET
HEIMARBEITSPLATZ
AUSGESTALTETER HEIMARBEITSPLATZ
ERGEBNIS STUDIE G
AUSWIRKUNG HEIMARBEITSPLATZ G
STUDIE AUSWAHL P
AUSWAHL EIGNUNG K
STUDIE EIGNUNG P
AUSWAHL AUSWIRKUNG K
AUSWAHL HEIMARBEITSPLATZ G
STUDIE AUSWIRKUNG P
EIGNUNG AUSWIRKUNG K
EIGNUNG HEIMARBEITSPLATZ G
```

The opportunity of precoordinating identifiers automatically is a noteworthy improvement of automatic indexing. It is exactly the assigment of complex identifiers that makes it possible for the searcher to formulate his query more precisely than with just a single identifier. What the linguistic treatment of text words is concerned automatic indexing has reached a level that can undoubtly compete with manual indexing. By means of inclusion of controlled vocabulary and thesauri into automatic indexing systems one can additionally put into the index as substitutes or complements the previously manually assigned identifiers.

## V. EVALUATION

In order to measure and compare the qualitiy of indexing systems one ought to perform representative retrieval tests. However these retrieval tests are quite expensive. Since out of reasons of time and costs such tests have not been possible, we have to restrict our comparison to an examplary one.

Let us again have a look at the several indexes of the introduced systems, especially a particular part of our example document, i.e. the title (Table 8).

At a cursory glance we realize that automatic indexing methods assign more identifiers as an indexer does who is used to a controlled vocabulary.

However, how convenient the search of our example document in a huge document pool is for a casual user who does not or hardly knows tne controlled vocabulary is dependant on the quality of the manual indexing of the contents of the title, and therefore is difficult to evaluate. In my point of view the problems of content in manual indexing become quite clear with this example.

Suppose, the searcher wants to retrieve this document he must reconsider the process of decision of the indexer, which means that he has to transfer the title, that he usually does not know, and then search with the help of this vovabulary.

The full text indexing usually assigns identifiers much closer to the text than manual indexing. These identifiers are much more convenient in the search strategie than the identifiers assigned by manual indexing. Nevertheless, indexing close to a text implies the problem that documents similar in content do not necessariely show the same linguisitc form. Thus, it is true, it becomes possible by means of truncation to dissolve linguistic differences which are caused by inflection.

Table 9: Evaluation

TITEL: Organisatorische Gestaltung von informationstechnisch gestützten Heimarbeitsplätzen für Programmierer (Teleprogrammierung)

| MANUAL INDEXING | FULL TEXT INDEXING | PASSAT | CTX |
|---|---|---|---|
| ARBEITSPLATZ-GESTALTUNG | GESTALTUNG | GESTALTUNG | GESTALTUNG |
| INFORMATIONS-TECHNOLOGIE | GESTUETZTEN | GESTALTEN | GESTUETZT |
| PROGRAMMIERUNG | HEIMARBEITSPLÄTZEN | STUETZEN | HEIMARBEITS-PLATZ |
| TELEKOMMUNI-KATION | INFORMATIONSTECH-NISCH | ORGANISA-TORISCH | ORGANISA-TORISCH |
| | ORGANISATORISCHE | PROGRAM-MIERER | PROGRAM-MIERER |
| | PROGRAMMIERER | | TELEPRO-GRAMMIERUNG |
| | TELEPROGRAMMIERUNG | HEIM | |
| | | ARBEIT | |
| | | ARBEITS-PLATZ | ORGANISA-TORISCHE GE-STALTUNG |
| | | HEIMARBEITS-PLATZ | GESTUETZ-TER HEIMAR-BEITSPLATZ |
| | | INFORMATION | |
| | | TECHNISCH | GESTALTUNG HEIMARBEITS-PLATZ G |
| | | INFORMA-TIONSTECH-NISCH | GESTALTUNG PROGRAMMIERER P |
| | | PROGRAMMIERUNG | |
| | | PROGRAMMIEREN | |
| | | TELEPROGRAM-MIERUNG | |



BEST COPY AVAILABLE

The questions asked in the beginnig shall now be answered:

1. What can be properly done by machine?

and

2. What remains to be done by man if indexing is performed by machine?

At present three different kinds of automatic indexing systems are available:

1. Inverted file full text indexing
2. Dictionary based indexing
3. Dictionary and rule based indexing

All three kinds of iondexing systems are used in pratice, full text indexing beeing the most frequent. This is the only method which is fully automatic. It is quick and cheap but dirty. The identifiers are text words in their original form.

With dictionary and rule based systems indexes are able to work more precise than in full text system. Inprocesing text words to get better indexes the quality is a function of the quality of the dictionary or/and the rule system, sophisticated text processing techniques and sophisticated dictionaries and rule systems. In short with increasing quality of the automatic indexing system the manual effort grows, which is necessary for controlling the dictionaries and rule systems. This effort remains the same, even over a long period of time, because language is continually developping i.e. new items have always to be added.

To come back to the questions asked in the beginnig:

HEIMARBEITSPALTZ' is transformed to a two-component-phrase 'GESTUETZTER HEIMARBEITSPLATZ', which makes no sense at all. To summarize: one may state that the automatic indexing methods available at the moment provide useful indexes, which are not worse than the manually produced indexes. Principally, these methods could be applied. One can only hope that, applying automatic indexing methods, the access to electronic stored information will be improved and thus an elevated supply (in information) satisfy the increasing demand of information.

Another point in favor of application of automatic indexing methods is the prospect of extending these methods to automatic translation.

## VI. AUTOMATIC INDEXING - THE FIRST STEP TOWARDS AUTOMATIC TRANSLATION

As has been explained, the demands on dictionaries and rule systems for text analysis get higher in keeping with increasing indexing quality. CTX has reached a level of analysis at which it is capable of automatically producing a translation concept by means of transfer-dictionaries and text generation methods. On the basis of CTX results an automatic translation system is developed in Saarbrücken which is supposed to make possible a rough translation from English into German and vice versa of great document amounts, even of whole databases (SHARMA, 1984). An application-oriented, userfriendly system is achieved, which ought to produce not a fully automatic high quality translation but an informative one. Therefore the emphasis is not on style but on the comprehensability of a text. The aim is to supply access for the international market to huge database of German.

The attempt shown in table 10 of translating part of our example document proves that a certain degree of maturity has already been

achieved. Single textphrases are translated rather suitablly (BEFASST MIT DEN ERGEBNISSEN EINER STUDIE - DEAL WITH THE RESULTS OF A STUDY). The system is already working pretty consistently, even if we have to acknowledge one or the other word is not yet contained in the English synthesis-dictionary (i.e. HEIM in HEIMARBEITSPLÄTZEN - HEIM WORKING PLACES). The results can be seen as rather satisfactory.

Table 10: Automatic Translation by ITS - Without Dictionary Improvement

Organisatorische Gestaltung von informationstechniech gestuetzten Heimarbeitsplaetzen fuer Programmierer (Teleprogrammierung)

ORGANISATIORAL ORGANISATION OF INFORMATION TECHNICALLY SUPPORTED HEIM WORKING PLACES FOR PROGRAMMERS (TELE PROGRAMMIND)

---

Befasst sich mit den Ergebnissen einer Studie ueber Aus-Auswahl, Eignung und Auswirkungen von informationstechnisch ausgestalteten Heimarbeitsplaetzen.

DEAL WITH THE RESULTS OF A STUDY OVER SELECTION, EIGNUNG AND EFFECTS OF INFORMATION TECHNICALLY AUSGESTALTETEN HEIM WORKING PLACES.



BEST COPY AVAILABLE- 

Like with all automatic indexing methods extensive automatic dictionaries are most essential. Correspondingly an exhaustive control of the dictionary is absolutely necessary.

Table 11 contains a sample of clauses which have been produced by ITS after revising the corresponding dictionaries. The ouput shows a rather good result.

Table 11: Automatic Translation by ITS - With Dictionary Improvement

| | |
|---|---|
| Verfahren zum automatischen Erkennen von Weissbloecken sowie Textbereichen, Graphikbereichen und-oder Graubildbereichen auf Druckvorlagen. | PROCEDURE FOR THE AUTOMATIC RECOGNITION OF WHITE BLOCKS AS WELL AS TEXT AREAS, GRAPHIC AREAS AND-OR GREY IMAGE AREAS ON PRINTED COPIES. |
| Verfahren und Anordnung zur Erzeugung von Daten und Eingabe der Daten in Computer. | PROCIDURE AND ARRANGEMENT FOR THE PRODUCTION OF DATA AND INPUT OF THE DATA IN COMPUTER.<br>PROCIDURE AND ARRANGEMENT FOR THE PRODUCTION OF DATA AND INPUT OF THE DATA INTO COMPUTER. |
| Pulscodemodulationssystem | PULSE CODE MODULATION SYSTEM |
| Magnetbandkassette | MAGNETIC TAPE CASSETTE |
| Geraet zum Aufnehmen und Wiedergeben von Tonaufzeichnungen | EQUIPMENT FOR THE RECORDING AND REPRODUCTION OF SOUND RECORDINGS |
| Verfahren zum Koordinieren einer Mehrzahl von Prozessen unter Verwendung eines Mikrorechners. | PROCIDURE FOR THE COORDINATING OF A LARGE NUMBER OF PROCESSES UNDER USE OF A MICROCOMPUTER. |
| Geraet zur Aufnahme und Wiedergabe elektronisch erfasster Daten | EQUIPMENT FOR THE RECEPTION AND REPRODUCTION OF DATA COVERED ELECTRONICALLY |

BEST COPY AVAILABLE

## VII. SUMMARY

The questions asked in the beginnig shall now be answered:

1. What can be properly done by machine?

and

2. What remains to be done by man if indexing is performed by machine?

At present three different kinds of automatic indexing systems are available:

1. Inverted file full text indexing
2. Dictionary based indexing
3. Dictionary and rule based indexing

All three kinds of indexing systems are used in practice, full text indexing being the most frequent. This is the only method which is fully automatic. It is quick and cheap but girty. The identifiers are text words in their original form.

With dictionary and rule based systems indexes are able to work more precise than in full text system. In procesing text words to get better indexes the quality is a function of the quality of the dictionary or/and the rule system, sophisticated text processing techniques and sophisticated dictionaries and rule systems. In short: with increasing quality of the automatic indexing system the manual effort grows, which is necessary for controlling the dictionaries and rule systems. This effort remains the same, even over a long period of time, because language is continually developping i.e. new items have always to be added.

To come back to the questions asked in the beginnig:

1. A machine is capable of automatically providing an index.

2. The better the index is supposed to be, the more the construction and the development of the dictionaries and rule systems becomes an effort.

Through the deployment of automatic indexing systems the indexer's work will change considerably. In the future he will rather work on terminology questions than assign identifiers. The concept of his work will have to be renewed, so that the traditional indexer will be substituted by a so-called 'terminologist'.

## VIII. LITERATURE

HOFFMANN, D., MÜSSLER, A., Automatische Selection von Stichwörtern aus Texten - PASSAT BS2000 Verfahrensbeschreibung. Müchen o.J.

KNORZ, G., Automatisches Indexieren als Erkennen abstrakter Objekte. Tübingen 1983.

LUSTIG, G., Weiterentwicklung der automatischen Indexierung im Projekt AIR. In: STROHL-GOEBEL, H. (Bearb.), 1984, S. 340-352.

PASSAT, Benutzerhandbuch. Müchen 1984.

SALTON, G. Automatic Indexing - A summary. In: TAYLOR, P.J., CRONIN, B. (Hrsg.), 1982 , S. 66-77.

SALTON, G., McGILL, M., Introduction to modern information retrieval. Auckland 1983.

SHARMA, K., Informative translation system. Unpublished paper. Saarbrücken 1984.

SONDERFORSCHUNGSBEREICH (Hrsg.): SALEM - Ein Verfahren zur automatischen Lemmatisierung deutscher Texte. Tübingen 1980.

SPARCK JONES, K., KAY, M., Linguistik in der Informationswissenschaft. Müchen 1976.

STEVENS, M.E., Automatic Indexing: A State-of-the-Art Report. Washington 1965.

STROHL-GOEBEL, H., (Bearb.), Deutscher Dokumentartag 1983. München 1984.

TAYLOR, P.J., CRONIN, P., (Hrsg.), Information mamagement research in europe. London 1983.

ZIMMERMANN, H. H., Automatic indexing and retrieval as a tool to improve information and technology transfer. Proceedings of the IInd scientific meeting: Computer Processing of Linguistic Data. Bled 1982.

ZIMMERMANN; H. H., u.a., CTX - Ein Verfahren zur computergestützen Inhaltsanalyse. Eggenstein-Leopoldshafen 1983.

IFLA GENERAL CONFERENCE, CHICAGO

1985

Division: Special Libraries

Section: Administrative Libraries

67-ADM-3-E

# LIBRARIES AND THE UNIVERSAL AVAILABILITY OF INFORMATION:

## Is Videotex the missing link in information services? - (international developments, examples and perspectives)

by

Gary L. Harris
Society for Information and Documentation
F r a n k f u r t / M a i n
FR Germany

*K E Y W O R D (S):*

Videotex; technology, service; trend; application; telesoftware; referral; user; information provider; retrieval.

*For internal use only:*

To be brought to meeting no:

162

Remarks:

to be presented orally ☒

for distribution only ☐



BEST COPY AVAILABLE

LIBRARIES AND THE UNIVERSAL AVAILABILITY OF INFORMATION:

Is Videotex the missing link in information services?

- (international developments, examples and perspectives)

ABSTRACT: An overview of videotex international developments is given. The special relationship of videotex to its increasing number of compatible technologies is discussed. Professional fields and types of information providers are looked at. Special service applications with focus on types, functionality and advantages are listed and described. Conclusions drawn are increasing international developmental activities and positive perspectives for successful application of Vtx to improve existing information services as well as to create a new generation of information services for both specialists and the general public.

LIBRARIES AND THE UNIVERSAL AVAILABILITY OF INFORMATION:

Is Videotex the missing link in information services?

- (international developments, examples and perspectives)

## Introduction

Since my last presentation of videotex in the library and Information field at the 49. IFLA conference in 1983, new trends can be recognized internationally. Worldwide, Vtx activities are increasing. Although the main concentration and most advanced development is and for the foreseeable future will remain in Europe, even small countries or city-states such a Malaysia, New Zealand, Indonesia, Kuwait, Brazil, Ecuador and Hong Kong are among those active with various applications. Another indication of the internationality of this medium is the in 1985 interlinkage of national systems in Europe. West Germany and France are the first two countries linked using the european "CEPT" standard. The West German PTT plans to continue to interlink its system with further european countries in the second half of 1985 and beyond.

## Vtx as the missing link of technology?

Vtx systems vary from centralized systems as in West Germany with its central PTT computer and storage to decentralized systems as in France without a central computer storage. The only common carrier in France is a telephone network - either public or packet switching to which individual information providers (IP's) are connected with main-frame, mini- and micro-computers. The same array of hardware is connected to the West German central computer extending the main service with external (independent) computers.

This array of hardware used and adapted for Vtx is of course by no means new to the library and information field. What is new is the simplicity and cheapness of Vtx technology, especially terminals (televisions) and the fact that this technology can not only serve internal needs well as an in-house system, but can also link the library or other IP's more cost effectively and efficiently to a much larger group of users, i. e. millions of homes and businesses, than otherwise possible. Whether it be banks for transactions and calculations, publishing companies for taking orders or online operators for information retrieval (IR), Vtx is becoming more and more the one universal compatible link to technology and services between producers and users of information in general.
Figure 1 shows the usual routes of accessing Vtx services.

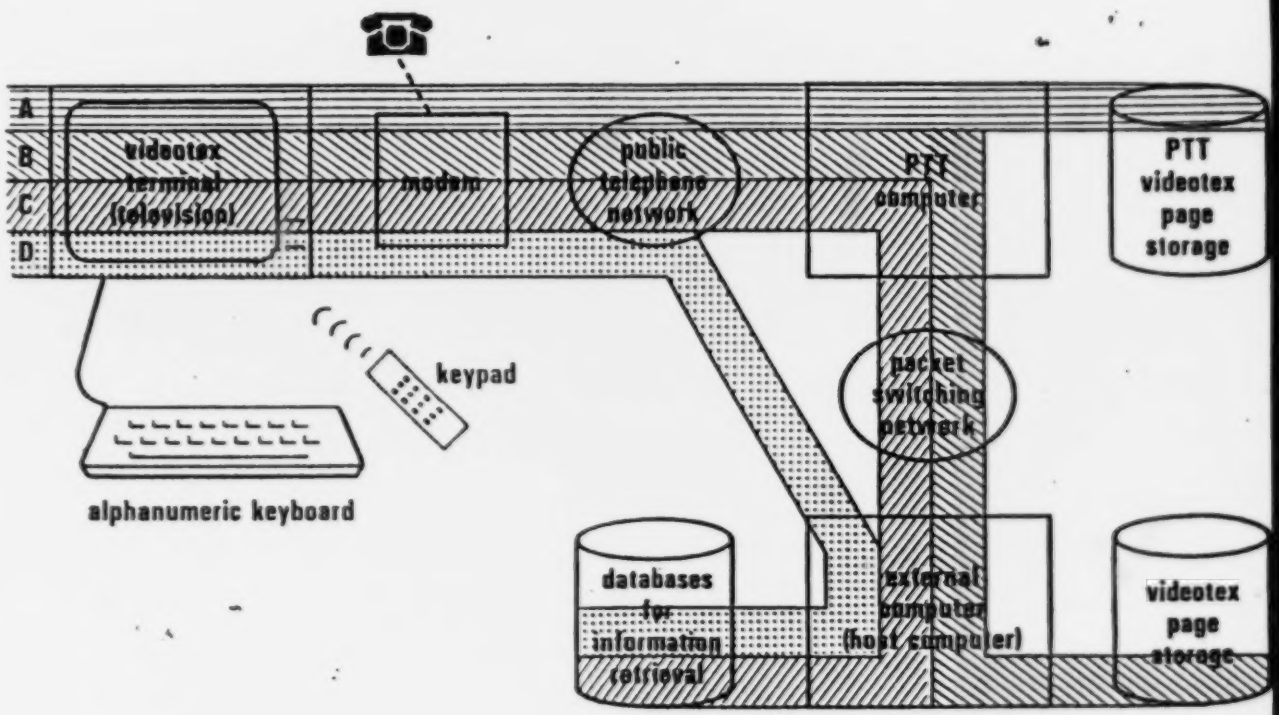
Who's interested?

Online operators have shown the most interest. Even a UNESCO Report has referred to the significance of Vtx to traditional online services. The interest is easily understandable. Databases which have been offered to special user groups can easily be made compatible with interfaces to the normal public Vtx systems. The database is consequently accessable by a potentially much larger user group than otherwise possible. Of course Vtx access doesn't replace the route of access used by a user who continuously searches in one special database. However, the occasional user who wishes to access many different databases has the advantage of only needing one cheap end-terminal and low cost public telephone network charges. Database providers are also implementing masking techniques so that users may select searching methods from menu selections without having to learn the different retrieval languages.

Administrative libraries as well as libraries which are more oriented toward a more general public are slow to adapt their computers to Vtx. This is due in part to their less favourable Budgets.

# Accessing Information with Interactive Videotex

The four routes of accessing stored information:




| | |
|---|---|
| A | videotex terminal to PTT computer with page storage for information programs |
| B | videotex terminal via PTT computer to external computer with page storage for information programs |
| C | videotex terminal via PTT computer to external computer with databases for information retrieval |
| D | videotex terminal direct to external computer with databases for information retrieval |

BEST COPY AVAILABLE



Figure 1 Accessing Information with Interactive Videotex

In spite of this, the number of libraries of different types i. e. libraries of the chambers of commerce in Germany are becoming increasingly active as Vtx IP's without opening access to their computers. This alone is a step, toward reaching out to the public. The mailbox and electronic ordering features are attractive features of the Vtx system and allow for better communication between libraries and between libraries and users.

Vtx has found interest in the educational field e.g. the Faculty of Continuing Education of the University of Calgary in Canada. Among other activities, microcomputers connected to Vtx for the handling of telesoftware enhance the possibilities of distant learning. University libraries could install Vtx connections as well as microcomputers in their self-instruction-centers to enhance services. Coursware could be developed and/or used cooperatively within a cooperative of organizations. Distant learning applications can even be cited in Indonesia and China.

The agriculture field has been increasingly utilizing microcomputers as well as Vtx for the transmission of telesoftware. The possibility should be closely examined by libraries to expand on already existing services using microfiche and AV-equipment to create program/ telesoftware libraries for such special user groups as farmers. Beyond the connection of microcomputers to Vtx is the use c video discs a new and developing application.

The medical library of a hospital in England is just one older example of activities in a number of countries applying Vtx to the medical field - in that case for the purpose of searching for medical literature. The West German Host "DIMDI" is also opening up its internationaly used medical databases to access over Vtx.

Some applications described in overview

Referral services can gain in quality when offered electronically over the Vtx system. Large indexes of institutions competent for almost every imagenable interest or information need are documented in printed form in virtually every library. If stored electronically they may be continuously kept up to date. Examples are information & documentation centers, information brokers, special libraries, professional associations, educational institutions, government offices, financial assistance and funding sources etc. Fast referral is possible using the easy menu searching technique as well as keyword searching. The user can either search directly or request information over the mailbox. In a large enough system such as in Germany and England, many sought sources will be represented in the same Vtx system as IP. This being the case, it is possible for the user to contact or access the referred to organizations immediately and therefore be aided more quickly and efficiently than otherwise possible.

Reference information of factual nature similar to addresses profits likewise by up-to-date and comprehensive storage. Examples are encyclopaedic works, dictionaries, telephone books, financial information i. e. stock market, medical dictionaries, pharmaceutical information, public documents etc. The Federal Statistical Office in the Federal Republic of Germany is an especially important developer of the accessability of factual data and factual databases over Vtx.

Bibliographic and copying i. e. document delivery services can be more comfortably used over Vtx. Experts in all fields in need of current literature such as from journal articles, conference proceedings etc. are often faced with significant delay in the documentation process before the needed documents are retrievable in online data bases or readable in printed services. Using Vtx frames, which can be updated within minutes, it is possible to provide subject oriented SDI-like references shortly after acquisition of any material. Upon viewing reference frames, the user can either aquire the desired article from his own library, in house literature collection, publisher or make use of the special convenience of the system by placing an other for copies on an interconnected response/orderframe upon which his/her Vtx account

is automatically billed with the from the IP determined fee. The user's account is simultaneously debited as the IP's account is credited by the PTT's system software.

In West Germany up to 9,99 DM or 3.40 US Dollars can be transfered automatically using such charge-frames. In France so called smart cards with individualized chips are being developed for billing purposes with Vtx.

An extension of the above mentioned services could be giving references and tips to subject related daily newspaper and magazine articles as well as radio and television programs. Book recommendations would be still another similar services.

Retrieval in traditional online data bases via Vtx gateway interface can either be an enhancement or alternative to Vtx system based services. It should'nt however be an either/or question for the IP because both services meet different user needs.

The simple interface to the Vtx system opens up access - as previously mentioned - to a larger user population, especially occasional users independent of end-terminals. It also requires the least attention as the IP does'nt have extra continuous work such as updating of Vtx frames. However, for user friendlyness, initial work should be done to mask searching commands and offer self-explanatory guidance in the searching methods and strategies.

For the user, direct access to the external data base should be seen as an enhancement to the services previously described offered on Vtx system frames.
The SDI-type of subject oriented literature references gives the user a look at the newest literature and can therefore meets his/her continuous information need. Specific problems or information needs are usually more occasional and consequently don't require continuous user access and precisely formulated searches. It should also be said that the best insurance of good queries or search results is when performed by experienced intermediaries.

Therefore the need of occassional precise searches will usually best be met by offering the user a mailbox or order-frame where he/she can indirectly type in a search request.

Calendars of events or announcements of the IP's own courses and seminars are attractive services. A random cross-section of events can be taken from a larger and more comprehensive online file and be offered on a limited number of Vtx frames, perhaps 5 to 10. This in itself can be a useful service and simultaneously give the user a good indication of the content of the entire online file. If more interest exists, it is always possible to subscribe to a regular printed service or according to occassional need place orders for print outs over Vtx which are made at the time of request and therefore contain all corrections or updates new since the last mailing of the printed service. IP's announcing their own courses, seminars, conferences etc. may offer the user the opportunity to register directly and immediately over Vtx, recieve equally quick confirmation, inform him/herself about cancellations, changes in program or fullfillment of registration limits faster than ever possible through the mail.

A special and in this case suitable enhancement to the described service would be the Vtx interface/gateway to the entire online file of calender events. Accessing a calender of events is quite different from accessing a literature data base. Above all, it is less complicated. The searching strategies are limited and can therefore make good use of making commands by allowing the user to fill out an electronic search form. Questions or even prompts may refer to subjects, location, name of organisor, date/months, price etc. The search is carried out automatically and should produce good results based on the simplicity of the content and cleverness of masking. The GID in Frankfurt, Germany used this technique in the old PRESTEL system and is further developing it for the new European CEPT standard.

Telesoftware or teleprograms can either be distributed over Vtx to users with compatible intelligent terminals or stored in central and external computer for ocasional etc. Possible applications are numerous.
Educational and agricultural teleprograms are significant in a number of countries. In the GID, one of many programs guide the user as a decision making tool to the special library which best meets his/her needs. Other programs may include searching aids, calculations of income tax, life insurance conditions, risk level of heart attack. Just as libraries keep documents, musical records and other optical as well as AV-media, the library of today and the future should seriously consider expanding to include an electronic teleprogram library for special user groups and general public.

## Conclusion

Vtx has reached a mature enough stage of general development around the would to provide a good basis of knowledge for analysis and decision making. The business community has shown the most interest and contributed the most to development. Experienced IP's or knowledgeable consultants in the library field aren't numerous. (Reports of activities would be appreciated by the author). Trends show Vtx becoming continuously more and more compatible with a variety of information technologies and consequently more universal in applicability. It serves not only to improve traditional services but also to create completely new services and channels of communication for both special user groups and the larger general public.

Literature

Bohse-Ziganke, T.: Bildschirmtext in Deutschen Industrie- und Handelskammern - Handelskammer-Bibliotheken auf dem Monitor.
In: Mitteilungen der Arbeitsgemeinschaft der Parlaments- und Behördenbibliotheken 55(1983), 5-16

Deider, C.: Ist der Einsatz von Bildschirmtext (Btx) in Bibliotheken denkbar?
In: Bibliotheksdienst 17(1983), No.2, 83-89

Ellis, B. ; Keenan, T.: Microcomputers, Videotex and Educational Teleconferencing.
In: Proceedings of the international conference on the application of mini- and microcomputers in information, documentation and libraries (Tel-Aviv 1983), 295-308

Harris, G.L.: Bildschirmtext in the Federal Republic of Germany: the role of the Society for Information and Documentation (GID) as an information provider.
In: Learned information: proceedings of the 7th International Online-Information Meeting (London 1983), 373-382

Harris, G.L.: Interactive Videotex application in the library and information field: a look at the developing West German scene.
In: INSPEL 1984, No.3, 214-226

Harris, G.L.: Providing specialized information via Videotex: german perspectives, strategy, program design, and gateways to Main-Frame and Personal Computers.
In: Proceedings of the INFODIAL-VIDEOTEX Conference, (Paris 1984), 145-149

Mathy, Armin J.: Das Pilotprojekt "Bildschirmtext-Bundesstatistik": 30 Monate Erfahrung in d. Anwendung von Bildschirmtext u. Schlußfolgerungen für d. Einsatzmöglichkeiten in d. Bundesverwaltung.
Wiesbaden: Statistisches Bundesamt, (ca.1981). -

Mittler, E. ; Wolf-Hauschild, R.:
Kooperation lohnt sich: Heidelberger Erfahrungen u. Überlegungen zur Zusammenarbeit öffentl. u. wissenschaftl. Bibliotheken auf lokaler u. regionaler Ebene.
In: Bibliotheksdienst 17(1983), No.1, 5-13

Norton, T.: Computerisation: a new dimension for library and information services.
In: Computerized systems in library and information services: proceedings of Aslib conference (1982), 3-13

# ADMINISTRATIVE LIBRARIES AND UNIVERSAL AVAILABILITY OF INFORMATION WITH SPECIAL REFERENCE TO INDIA

By

M.K. JAIN*

## ROLE OF ADMINISTRATIVE LIBRARIES:

Administrative or Government Libraries are those whose primary purpose is to serve the nation at a Government level. These libraries at the national level are set up in various Ministries and Departments to serve the legislative, administrative, and developmental functions of the Government. These libraries mostly collect official gazettes, parliamentary debates, census reports, gazetteers, annual reports, committee and commission reports, manuals, budget papers and acts, serials, monographs, and other published material in the areas of their specialization.

Information today is an essential input in all activities, because of its value, relevance and importance for decision-making, research, and planning. All organizations whether government or private, use information in every aspect of their work for optimal utilization of resources to achieve maximum out-put. National development and the process of decision-making depend upon the quality of information made available to research analysis and for structuring information content. Thus the role of administrative libraries has changed tremendously than what it used to be due to the assumption of diversified developmental activities. Now the librarians working in

* Chief Librarian-cum-Documentation Officer, Planning Commission, Yojna Bhavan, New Delhi
President, Government of India Librarians Association,
Vice President, Indian Library Association.

the governmental set up have to identify and retrieve relevant material from the information generated in the form of published books and monographs and also from non-book sources. This information is furnished to the administrators, scholars, and researchers without their asking. Because of this important role their number is increasing day by day. The total number of administrative libraries in the four countries viz. Canada, INDIA, U.K., and U.S.A. is given below:

| | COUNTRY | NO. OF LIBRARIES |
|---|---|---|
| 1. | Canada | 292 |
| 2. | India | 6,000 |
| 3. | U.K. | 905 |
| 4. | U.S.A. | 1,591 |

ADMINISTRATIVE LIBRARIES IN INDIA

It will thus be seen that India has got the largest number of administrative libraries in the world. The organizational set up of the Government of India consists of various ministries which are further divided into various departments, divisions, branches, and sections, besides their attached and subordinate offices. Section is the smallest administrative unit which deals with the specific subject in a Ministry/Department. The library of the Ministry/Department is generally regarded as section in the organization. With the exception of a few libraries which have the status of a Division or a Branch, other libraries constitute smallest units of the organization or sometimes as part of the section.

Real progress in the expansion and updating of administrative libraries in India has taken place after independence and that too after the planning era in 1950s. The administrative libraries in India, therefore, exist for the fulfilment of the aims and objects of their parent ministries and departments. As a result there is hardly any area or discipline which has not been covered by one or the other Government departmental library. To illustrate this point, grouping of a few administrative libraries, having common subjects of interest is given at Annexure I.

GOVERNMENT PUBLICATIONS

Government publications are so important today that their number is increasing day by day. The Government of India and the Planning Commission were seized of the gigantic task of restructuring the national economy for the socio-economic uplift of the masses. In this context, the need and scope of government publications increased considerably and assumed new dimensions. The production, distribution, and sale of publications was broad-based with the twin objectives of informing the people about the policies and programmes of the Government and to provide useful information/data to all segments of the society viz. farmers, academicians, scientists, administrators, technocrats, industrialists, student and scholars etc. for research and planning. Government publications are not only important in their country but also needed by other countries. The Government publications brought out in UK, USA, Canada, and Australia are very much in demand in India. They are acquired through various exchange arrangements, and sometimes

purchased also. Similarly other countries viz. UK, USA etc. acquire publications of the Government of India under such arrangements. The Controller of Publications, under the Ministry of Works and Housing*, brings out publications on behalf of the various Ministries and Departments. This agency is just like HMSO, London in UK.

Central Sectt. Library is one of the recipient libraries under U.S. Government exchange programme which receives all US Government publications. Previously it was getting all the publications from the main coordinating agency i.e., Smithsonian Institute, Washington. It is now getting all Government publications on micro-fiche from the Library of Congress, Washington. Besides the Central Sectt. Library, two other libraries i.e. National Library, Calcutta, and Parliament Library, New Delhi, have also made exchange arrangements for getting UK and US Govt. publications. These libraries are obtaining government publications from Canada, Australia, New Zealand, etc. from various sources and sometimes make purchase also. Experience has proved that the present arrangements for getting Government publications is not satisfactory. The problem can be solved if in the overall strategy of UBC and UAP, all the Government publications are also included. There is need for making universal availability of information in administrative libraries of all the countries. This can be done when there is proper bibliographical control in each country and the publications are made available universally without any constraint. Perhaps, comparative research for evolving a reasonably acceptable classification if not a standardized system could be undertaken by IFLA.

---

*Dapartment of Publications, Civil Lines, Delhi.

UNIVERSAL AVAILABILITY OF INFORMATION

IFLA's programme of Universal Bibliographic and UAP are well known by this time. UAP is supported by UNESCO and has been the subject of an International Congress. The IFLA/UNESCO International Congress on UAP held in Paris in May, 1982 was aimed at Librarians and other Information personnel, and also at policy makers in government and elsewhere. Since UAP is essentially a programme for action, the Congress tried to demonstrate the need for action and indicate the possible course of action.

The objective of UAP programme is to promote and increase the availability of publications. To achieve this, several measures are needed. Relevant sectors of the community and governments must be made aware about the programme of IFLA and should also be motivated to acquire the information generated. For a convincing case to be made, publicity is not enough; both the importance of document supply and the deficiencies in the present situation need to be highlighted. A good case could be made that modern society, and future society even more, are dependent on information for survival and advancement. Publications are not only the vehicle of information, but they are a main vehicle and are likely to remain so, whatever the technology changes might be in the form and nature of publication.

Deficiencies in the present situation are a matter of every day experience to all concerned with providing and using library and information services. To argue a need for availability and to show that present situation is unsatisfactory will not of themselves help

to improve the position. Some guidance is, therefore, required as to how problems might be solved, not only in developing, but in some developed countries, where the resources provided to libraries are diminishing in real terms year by year, possible solutions should be offered taking the emerging needs and the availability of resources.

Since each country and library needs access to the holdings of libraries in other countries, each of the above factors must be viewed locally, nationally, and internationally.

BIBLIOGRAPHIC CONTROL IN INDIA

Availability of information at a global level is possible only when there are regional bibliographical centres throughout the world. Effective functioning of regional bibliographical centres is possible only when the countries in that particular region have got their own bibliographical control. In India a number of institutions/agencies are engaged in bibliographical services relating to national output in the fields of their concern. It is not possible in a big country like India that a single institution can have a comprehensive bibliographical control of the total output. In order to have a total bibliographic control, efforts of many institutions and agencies may be shared and coordinated according to a well conceived plan of action. At present the following institutions are engaged in the bibliographic control of Indian sources of information:

| | Institution | Title of Publication | Periodicity |
|---|---|---|---|
| 1. | Central Reference Library, Calcutta | Indian National Bibliography | Monthly, Annually Two-yearly and five-yearly cumulation. |
| 2. | Registrar of Newspaper, Ministry of Information & Broadcasting, N. Delhi | Press in India (two parts) | Annual |
| 3. | INSDOC, New Delhi | Indian Science Abstracts | Monthly |
| 4. | Association of Indian Universities, N. Delhi | University News | Three times a year |
| 5. | Social Science Documentation Centre, Indian Council of Social Science Research, New Dehli | Union Catalogues of Social Science Periodicals, Serials, Newspapers in Delhi and other States of India. Among other publications there is a SSDC Research Information Series. | No specific periodicity |
| 6. | Controller of Publications, Dept. of Publications, Ministry of Works & Housing, Civil Lines, Delhi-54 | Catalogue: Government of India, Civil Publications | Monthly & Annually |

A national seminar on Bibliographic control was held in India at the National Library, Calcutta in February 1985. The following proposals were made:

1) Adoption of a methodology for having a standardized processing system for National Bibliographic Control;

2) Modifications of different acts to enable proper flow of books to the National Library for the preparation of the Indian National Bibliography;

3) To evolve the scope and coverage of INB in order to have better bibliographical control;

4) Methodology for standardized retrospective bibliographies and union catalogue;

5) Feasibility of standardized automated bibliographic system for the INB and National Library in conjunction with other large collections in India.

At present several organizations in India have been experimenting in the use of computer in handling information. Some of the organizations which have been involved in these activities are (1) Bhabha Atomic Research Centre (BARC), Bombay; (2) Tata Institute of Fundamental Research (TIFR), Bombay; (3) National Informatics Centre (NIC), New Delhi; (4) Documentation Research and Training Centre (DRTC), Bangalore; (5) Indian National Scientific Documentation Center (INSDOC), New Delhi; (6) Indian Institute of Technology, Madras; (7) Bharat HeavyElectricals Limited (BHEL), Hyderabad.

INSDOC has used computers in jobs like preparation of indexes to various publications, processing of union catalogues, data compilation, inventory of serial publication, preparing directories, etc.

BARC, TIFR, and IIT(M) have also used computers for certain library operations. BARC has been experimenting the SDI services with INIS data base, whereas DRTC has developed programmes for SDI services using ICL 1901 computers. This system has been applied in the SDI services being offered to the users in the BHEL Complex. DRTC has also developed a system called DOCFINDER.

Large scale computer application in information field has been made in late 70's with the introduction of SDI service by INSDOC in collaboration with the IIT Madras. The project was supported by UNISIST and was based on CAN/SDI software. SDI services were offered against 130 profile users from India, Nepal, and Sri Lanka.

National Information Centre (NIC) at Electronics Commission (IPAG), New Delhi developed an integrated online bibliographic database for library documents using CDC Cyber 170/720 main frame computer system.

NIC's Library Data Base accessed by NICNET (NIC Computer Network) members connects to Cyber System.

CONCLUSION

The above account is not comprehensive of all the problems which need consideration in the matter of bibliographic control, organization, and dissemination of Indian Information Sources. The important points which have the bearing on the introduction of information technology have, however, been mentioned. It will be thus seen that in India the bibliographic control as compared to other countries is quite adequate. During the Seventh Five Year Plan i.e. 1985-90 it is expected that Indian National Bibliography in our country will provide computer based network. Thus in due course of time India will be in a position to participate effectively in the activities of the global network of Universal Availability of Information. Similarly the other countries may develop infrastructure for network of their national bibliographic control so that regional centres when opened, function smoothly and efficiently in the overall strategy of UAI.

REFERENCES

1. Bowker Annual, 1984.

2. Government Libraries in India and their role in Dissemination of Information - a lecture by M.K. Jain at National Medical Library, New Delhi on 7th May 1985.

3. Guide to Government Department and other libraries 1984 London, British Library, 1984. 104 p.

4. Handbook of libraries, archives and information centres in India, New Delhi Information Industry Publications 1984-85, Vol. I, Libraries and Archives.

5. Papers at the National Seminar on Bibliographic Control in India, 25-26 February 1985 held at National Library, Calcutta, India.

6. Report of the Working Group on Modernization of library services and informatics for the Seventh Five Year Plan 1985-90. New Delhi, Planning Commission 1984. 95 p.

ANNEXURE I

HISTORY, SOCIOLOGY AND CULTURE: The libraries of National Archives of India, Archaeological Survey of India, and Anthropological Survey of India are grouped under this subject-specialization.

ECONOMICS: Libraries of Ministries of Labour and Employment, Finance, Industry and Commerce, Planning Commission, Department of Revenue and Insurance, Directorate of Economics and Statistics, and Department of Economic Affairs have comprehensive collections on economics and allied subject.

PUBLIC ADMINISTRATION: The collections in the libraries of Ministry of Home Affairs, Central Bureau of Intelligence, Ministry of External Affairs have a wide variety of publications on this multidisciplinary topic.

SCIENCE AND TECHNOLOGY: In contrast to the social sciences, humanities, art and culture collections, where the richness of literature is measured through standard works, the emphasis in science and technology is placed on acquisition of latest publications and periodicals. Many libraries, covering both pure and applied sciences, such as medicine, agriculture, engineering, etc. have emerged and developed after the independence. The libraries of the Ministry of Defence, Agriculture, Department of Science and Technology, Director-General of Health Services, Central Board of Irrigation and Power, Central Water and Power Commission, etc. have useful specialized collections on their respective areas and disciplines.

COMMUNICATIONS AND TRANSPORT: The libraries of All-India Radio, Posts and Telegraphs, Press Information Bureau, Ministry of Information and Broadcasting, Ministry of Railways, Department of Civil Aviation, etc. have specialized collections in these disciplines.

BUILDING AND ARCHITECTURE: The libraries of Ministry of Works and Housing, National Building Organization, Central Public Works Department, etc. maintain collections on town planning, architecture, environmental studies, etc.

SURVEY LIBRARIES: Surveys have been playing an extremely valuable role in our national developmental process. These are the oldest organizations established to study and identify natural, human and cultural resources of the country. Archaeological Survey of India is engaged in maintaining our historical monuments and sites, and also bringing into light many aspects of our ancient civilization by excavation in different parts of the country. The Anthropological Survey of India is engaged in the study of cultural and physical apsects of different communities in the country. Botanical Survey of India and Zoological Survey of India are responsible for the preservation and development of our flora and fauna. Similarly, the Geological Survey of India has been busy in identifying and exploring our natural resources. There is a network of libraries maintained by these all over the country and surveys cover publications on subjects such as natural sciences, social sciences, arts, architecture, engineering, history, etc. These libraries have very rich collections on subjects in their respective specializes areas.

MUSEUMS: The collections on museology, fine arts, culture, architecture, are maintained by the Indian Museums, National Museum, National Gellery of Modern Art and the State owned museums.

LEGISLATION AND LAW: The libraries of Parliament, Ministry of Law, Supreme Court, etc. have large and valuable collections of literature on legal and parliamentary affairs of several countries.

IFLA GENERAL CONFERENCE; CHICAGO 1985

Division: Special Libraries

Section: Administrative Libraries

1985

85-ADM-4-E

# Literature data banks for institutional libraries

## A critical comparison between conventional information media and the possibilities offered by EDP-supported legal information systems in Europe

by

**Gerhard Kruse**
European Patent Office
Erhardtstraße 27
D-8000 Munich 2
Federal Republic of Germany

For internal use only
to be brought to meeting no:
.....162.....

Remarks:
to be presented orally ☒
for distribution only ☐

BEST COPY AVAILABLE

The number of laws and law-related technical and economic rules and regulations is growing year by year. The associated literature, commentaries, textbooks, monographs and periodical articles are becoming more and more extensive. The number of final-instance legal judgments, which are constantly required for scientific and practical work, can be recorded bibliographically only with the greatest difficulty, since only a relatively small, often subjective selection is ever printed and thereby publicised.

In the normal fields of law, such as civil, criminal and constitutional law, and even in public law, the collections of legal texts and commentaries are usually well-maintained. They are still manageable for the lawyer and librarian at the national level. The difficulties begin with special areas of law, particularly when regional and local law is involved. Who, as a librarian, can be sure that a legal text in a loose-leaf collection or in a commentary several years old is still valid law? And who but the librarian should be responsible for up-to-date information? Perhaps the commentator has died or the publishers, for commercial reasons, have not brought out a new edition. It is then inevitably necessary to

read through legal gazettes, even though it is not clear whether the correct legal gazette is being read through. We are all familiar with the user's comment: "I am quite sure there must be something more recent!" It is then very difficult to prove that something does not exist! A librarian at an institutional library, by its very nature lacking the reference books of large university libraries, often cannot help the user in such matters. We all know this vicious circle. An important decision cannot be traced bibliographically because it has been omitted. An escape-goat has to be found; the library is no good! During the next budget negotiations the financial allocation is cut. Perhaps one day it is restricted to personal reference libraries. This would militate against a system of unrestricted dissemination of information aimed at facilitating the user's task and providing him with all necessary information sufficiently quickly for the task to be completed as economically, successfully and authoritatively as possible.

As a result of increasing international collaboration, e.g. in world-wide trade and common defence efforts, access to the laws, regulations and even very specialised texts of other countries is required by users more and more frequently.

This requirement for information can be illustrated by referring to two occasions when I myself had to find the relevent legal provisions. An entrepreneur wanted to know the maximum carbon content in lorry brake pipes permissible under United States statutory regulations. On another occasion a manufacturer of carrier-frequency equipment wanted to know the level of cross-talk attenuation still tolerated on cables in Chile.

One could say of course that all that is not the librarian's task. But who is supposed to obtain such information if the librarian merely shrugs his shoulders?

The need for information is very extensive: every district court, local authority, legal department of a small industrial concern and every lawyer constantly requires such information. However, very often he does without it because he thinks it is difficult, or because it really is difficult, to obtain the information and documents. It is this resignation on the part of many users in the face of the difficulty of using bibliographic aids, often because they do not themselves possess them, which leads to superficial work.

However, the most serious problems arising from insufficient bibliographic search in the field of law occur when laws are amended and regulations issued even though they are in direct contradiction to existing statutes. Such regulations often affect unforeseen aspects of life. Above all, legislators do not make use of the experience of others who have had to cope with the same or similar problems in the past, either out of arrogance, or lack of consideration. The overall state of legal documentation represents a challenge to information science.

A review of the present state of sources in Western Europe shows astonishingly that over 3 000 law gazettes and official journals are published regularly at the regional level of département, Bundesland, etc. In addition around 500 official journals, etc. are published by government ministries.

The National Prussian Heritage Library (Staatsbibliothek Preußischer Kulturbesitz) collects the official publications of around 30 States, but normal periodical lists are seldom included, usually only special sections of national bibliographies containing references to official publications. It is not the purpose of this paper to report on official documents, their form of publication and the bibliographical references to them, but I must point out the difficulties arising even in recording the extensive sources. Classification is relatively simple since either geographic and political boundaries are followed or particular legislative responsibilites are clearly defined. When capturing secondary literature, however, particularly comparative legal works, problems arise immediately.

The problem of analytically capturing legal literature has in no way been solved. However, the user is more dependent on clear and comprehensive information than in other subjects, if he does not want to be at the mercy of lawyers who, as he does not discover until later, are themselves less well-informed than the keen library user.

It may have something to do with the lawyer's approach to his work. Unlike the scientist, he does not refer to all relevant documents, but relies on particular arguments which appear to him to be suitable and convincing, thus simply withholding from the reader contrary opinions and arguments. It is then up to the reader himself to draw up a bibliography and study the literature, if he wants the results of his study to be objective and really scientifically-based. However, as librarians we are committed to scientific working methods, guaranteeing balanced arguments and clear logical reasoning. Despite attempts to create comprehensive running legal documentation with satisfactory analytical capture, including peripheral subjects, it has obviously been unavoidable, as a result of the number of titles, that papers and judgments have to be limited to a selection. Thus only about 15 000 works per annum are bibliographically referenced in the Federal Republic of Germany for all fields of law: works that are also frequently cited in the relevant commentaries to laws. In view of the material, which can only be articulated by means of language, this is extraordinarily little, remembering that nearly half a million people are employed in legal professions in the Federal Republic. The volume of material to be processed is clear from the number of laws and regulations in the same field

About 12 000 federal or regional laws and regulations enter into force or are amended every year. In addition there are about 6 000 administrative regulations and 2.2 million court judgments, half of which represent Supreme Court jurisdiction. About 5 000 monographs are published annually in the field of jurisprudence alone, plus an equal number of dissertations and about 300 new editions of commentaries.

Specialised legal bibliographies are published in nearly all European countries, some retrospectively, others on a running basis, but frequently periodical articles are not cited. Even European Community documentation has only been bibliographically captured since October 1984. The present situation is totally unsatisfactory for the librarian who has to provide his user with bibliographic data. The by now familiar technical and organisational opportunities provided by data processing could radically alter this unsatisfactory situation. Access to laws and regulations and even to legally relevant, technical standard works would be possible even from the smallest institutional libraries. Articles and relevant commentaries could be included in full with the judgment and not, as hitherto, only those arguments deemed suitable by authors

of legal works. Librarians must resist over-specialisation. To simplify usability, it would be preferable to have a system covering all fields of jurisprudence. The structure of such a system could be broken down as follows:

1. Laws, regulations and rules with legal force

2. Collections of decisions and court judgments

3. Legal data files (e.g. patent register, trademark register, penal register etc.)

4. Literature on particular fields of law, legal opinions.

It will not always be possible to use the methods of capturing customary for literature data banks. For certain fields classifications may be necessary to reduce redundancy and guide the user.

Three basic questions have to be asked at the outset of such a project:

1. Who will organise the appropriate data banks without rep
senting in any way economic, political and ideological
interests and arrange digitisation of all relevant docu
ments, including peripheral subjects? In the field of
unlike chemistry for example, this work cannot be giv
a private company.

2. Who will carry out the analytical capture according
standard criteria compatible with the systems of o'
countries, in the event that documents have to be
internationally?

3. Who will cover the cost of such a project until
self-financing, as I am sure it will one day be

All countries spend a lot of their tax revenue on
library books and periodicals, but only in excep
do they invest in analytical capture. Only a ve
portion of monographs are included in national
It became no longer possible to cope with the
mentation using traditional library methods o

BEST COPY AVAILABLE

material, since cumulative collections cannot be printed out in paper form, yet card catalogues can only be set up in a few places for reasons of cost. As a result comprehensive capture of such information was neglected even when documentation in the field of law was thoroughly desirable.

Unfortunately it has not been objective requirements which have forced users of information in the field of law to act, but the manufacturers of data processing systems wanting to sell their products who have pressed for the introduction of new technology. Programs and equipment have been marketed often without any knowledge of the volume of material involved and the problems of capturing it. It is easy to get approval for funds for projects involving new technology, the final scope of which cannot easily be foreseen. A typical example is the EURODOCDEL data bank intended to record European Community documentation. The documents will be captured not by trained permanent staff, but by temporary staff recruited under job creation schemes. To ensure that the system is well-distributed, searches are to be carried out via videotex according to French BTX standards. Conversion to other countries' systems

is not technically difficult, but cost-intensive. Response times, already very long, are doubled, the number of letters per line is limited and the number of lines itself is small. Concessions have been made to encourage potential users. The project was subsidised at a cost of around $ 9 million, without a preliminary market analysis being carried out to find out the actual number of potential users. Integration into an overall system would have considerably reduced equipment costs and the analytical capture of documents would only need to have been carried out once at source.

The documents themselves were stored on video disk and can be printed out via laser printer. They can also be transmitted by telecopy from Paris, although high telephone charges are incurred (3 minutes per page).

It would have been cheaper to re-organise the entire word processing systems of European institutions, establishing norms for word processing so that documents are already in digitised form. A cheaper form of transmission, e.g. via Datex-P or teleprinter would thereby be possible. The technical facilities for this now extend to every postal network in Europe. The EUROLEX data bank, operated by an affiliate of the International Thompson Organisation, is also accessible via Datex-P and thus throughout Europe. The full text is

stored, the documents being captured by automatic indexing. Thus, for example, the decisions of the Boards of Appeal of the European Patent Office are recorded by EUROLEX. Indeed, court decisions form the bulk of the data collection. Legal regulations are also being increasingly incorporated. EUROLEX is geared towards information centres, such as law and institutional libraries, not so much towards the individual customer. Now a few words about full-text storage. This is an attempt to overcome the difficulties of conscientious analytical capture by using technical methods. The author's expressions and technical language are retained and it is up to the researcher to use the same language during his research. Experience shows that the amount of exploitable information thus obtained, depending on the subject, is considerably less than that provided by a search carried out using well-defined keywords. In addition, the intellectual effort, required only once when documents are analytically captured, has to be repeated every time in the case of a full-text search.

However, for the librarian intent on objective information it is disconcerting to find a private company which publishes daily newspapers, books and magazines and is also active in the oil and tourist business selecting the documents to be searched by a user concerned with scientific accuracy.

There are a large number of data banks like EUROLEX covering specialised fields of knowledge. The demarcation between them is still in a state of flux, as are the cases where the recording of bibliographic data is duplicated. Market forces are followed and studies made on which users have the most money to spend on acquiring a data bank. In the days before data banks a specialist would draw up a bibliography for a particular field, e.g. law, which would then be printed. For a library to acquire it the cost would be $ 100 at most. Researching a data bank generally costs more. Over $ 1 000 are necessary for the terminal alone. Every search costs money. The psychological pressure should not be overlooked. To obtain a password for a data bank it is often necessary to invest considerable amounts of money. What user with a budget often insufficient for acquiring the most important equipment and literature is going to risk expensive searches for which he does not know the exact cost in advance nor how far they may

extend? And if he cannot say at the end: "There is nothing else, I have researched carefully, I can be quite sure", he may abandon the search altogether and formulate his work according to his own ideas. For that we need no libraries or trained librarians. The same user would no doubt have spent a number of hours looking through bibliographical reference works and would often have come across a good deal of other relevant information. In addition to bibliographical studies, the amount to be gained incidentally, consciously or unconsciously, from reading through titles, even without reading the articles themselves, should not be under-estimated.

The marketing of information has also become more expensive. Whereas in the past only printing costs were incurred, nowadays there are the costs both for digitisation and for continuous availability on electronic storage media. It is no longer worth issuing printed editions to be placed on shelves, since the bibliographic data are sold to a data bank. And it is inherent in the system of data banks that I can no longer know what percentage of published literature I have in fact recorded or whether I have perhaps been searching in peripheral fields using a completely incorrect search strategy; perhaps I have been trying to get my information from a data bank which is very weak in that particular field.

It can be very disappointing to discover how much a host really has to offer, despite a big advertising campaign. I say this in a deliberately provocative way, but business interests based on the profit motive are simply incompatible with serious bibliographical work aiming at scientific comprehensiveness.

The German data bank JURIS, operated as an independent institution, in which all higher-instance German case law including literature since 1977 is stored represents a step in the right direction, leaving aside the GOLEM retrieval system. GOLEM is not compatible with other data bank systems. But large-scale projects also have their problems. JURIS is therefore so far unfortunately only accessible to a very limited circle of users.

To present here existing data banks in their entirety with their weaknesses and advantages would go beyond the scope of this talk. For this purpose there are data bank guides as well as extensive literature which I shall list at the end. The most important data banks, however, are:

LAWTEL
CEDIJ
ITALGIURE

Others: Patent data banks (INPADOC)
Penal register data banks, data banks for technical rules, literature data bank on criminal technology, DATEV-LEXINFORM (a data bank for tax consultants)

But lectures are supposed to stimulate discussions. There is clearly a great need for legal information, but we have not succeeded, as in nearly all fields of documentation and information, in setting up a system which, apart from recording all peripheral fields, offers maximum comprehensiveness, makes international searches possible with a standard search strategy and provides the user with reliable information on whether a law or regulation, issued decades ago, is still in force and clashes with other laws. It would also be useful to know if any legal decisions have been taken recently with respect to that particular regulation or whether courts have ceased to apply a law altogether.

I am aware that little can be done to counter the keen economic interests of producers of data processing systems with a lead in programming or of those trading in bibliographical information. Many of these current projects are already doomed to failure because of their clumsy structure and the consequent small number of users. It would then be time for the IFLA to intervene in the hitherto haphazard trend with clear-cut recommendations, based on specialised librarian knowledge and experience, the failed companies being taken over by independent institutions. It would be desirable for the Administrative Libraries section to find a worthwhile task in this field over the next few years, so that data processing specialists and self-styled documentalists do not simplify the problems by themselves in their "traditional" way.

Bibliographical References


Drotnek, Susanne: Bibliotheken als zukünftige Anbieter und Benutzer von Bildschirmtext. Hannover-Waldhausen: Nordwestverlag 1984. - 102 Schrift. z. Bibliothekspolitik Bd. 2

EDV und Recht. Möglichkeiten und Probleme v. F.J. Bauer u. a. ; Hrsg. Arthur Kaufmann Berlin: J. Schneider 1972

Gesetzesplanung. Beiträge zur Rechtsinformatik. Hrsg. v. d. Arbeitsgemeinschaft Rechtsinformatik München - Regensburg. Berlin: Schweitzer 1972

Haft, Fritjof: Elektronische Datenverarbeitung im Recht. Berlin: J. Schweitzer 1976

GELN Datenbankführer. Dialogfähige Datenbanken der deutschen DIANE - Kosta [GMD, FIZ Technik, GID und INKA. Stand März 1983] Frankfurt/M: DV Verl. f. internat. Dokumentation Werner Flach K. G.

Frastal, Bernhard M. : Datenverarbeitung im Bereich juristischer Dokumentation. Ein Arbeits- und Funktionsvergleich zweier Systeme. Berlin: Schweitzer 1975

Frohnde, Werner: Users of Legal Information Systems in Europe. München: Schweitzer 1981

Wallberger, Otto: Studien zur Normenlogik und Rechtsinformatik. Berlin: Schweitzer 1974